Simulation of the formation of clusters of bioconjugated FeO nanoparticles after their specific interaction with a biomarker

Pablo Palacios Alonso

University Authonome of Madrid
pablo.palaciosa@estudiante.uam.es

November 27, 2020

Abstract

In this work, we have build a computational model to perform a comprehensive study of the formation of clusters of nanoparticles functionalized with a ligand which interacs specifically with a biomarker. In our model the interaction between the nanoparticles and the biomarkers is always mediated by the ligands and the only interaction between the nanoparticles is a Lennard-Jonnes force that prevents them from overlapping. The influence of the concentration of nanoparticles and biomarkers and the number of ligands per nanoparticle in the mean size, compactness and dispersion of sizes of the clusters has been studied using several combinations of parameters. The results have been compared with experimental mesures to check the validity of our model and we have find out in which ranges of parameters our model is valid and in which ranges there are non-negligible interactions that we are not taking into account.

1 Introduction

In recent years, nanotechnology has proven to have a huge number of medical applications. Drug delivery [**?**], selective destruction of tumors [**?**], or biosensing [**?**] already benefit from using nanomaterials to improve the current healthcare standards. Nowadays, the most widely used techniques to detect biomarkers are immunoassays which have some limitations related to the need for an exhaustive sample preparation, to involve complex biochemical processes, and to the limited capacity to modulate detection sensitivity. The use of magnetic nanoparticles (MNP) has emerged as a promising solution to solve all this limitations. Recent results show that different techniques can be used to detect the presence of biomarkers using some changes in the magnetic properties of the MNP after the interactions with the biomarkers [**?**]. Among the used techniques, the magnetometry AC provides high sensibility in the detection of biomarkers (to 1 nM), do not require an exhaustive preparation nor long incubation times (<30 min) and the visualization of the detection can be made with commercial systems easy to use [**?**].

Magnetometry AC uses the change on the magnetic relaxation mechanisms of the MNP when they form clusters after the interactions with biomarkers (analytes) as a way of detecting its presence. The interaction between MNPs and biomarkers is mediated by a ligand capable of interacting with the biomarker. That ligand is linked to the MNP in a process called bioconjugation. The formation of clusters takes place when the valence of the biomarker is equal or higher than two, in this case, two or more MNP can be linked to the same analyte forming bonds between the MNP and then the aggregates (Figure **??**). The change on the magnetic relaxation mechanisms depends on the size and shape of the clusters, that's why it's essential to understand how varies both magnitudes with the different variables of the problem: valence of the biomarker, concentration of MNP and analytes and the number of ligands per MNP.

Figure 1: *Scheme of several MNP, some of them joined by analytes*

In our work, variation of the size and the shape of the clusters when the valence of the biomarker is equal to two and the rest of the parameters vary has been studied. The analysis has been made using an algorithm of Brownian Dynamics (BD) which is a method used to solve the motion of colloids of sizes between $\sim$ 10 nm and $\sim$ 1 μm immerses in a fluid. On that length scale collisions between molecules of the fluid

and colloids can be substituted by a stochastic force and friction force is big enough to neglect the inertia of the MNP. Under this approximations, the colloid's equation of motion is [**?**]:

$$\mathbf{dr} = \frac{\mathbf{F}}{\xi} + \mathbf{d\tilde{r}} \tag{1}$$

Where ξ is the friction coefficient, $\mathbf{F}$ the external forces that acts on the particle and $\mathbf{d\tilde{r}}$ is a random motion that collects the noise force contribution and, verifies $\left\langle \mathbf{dr}^2 \right\rangle = 6D_0 dt$ where $D_0 = \frac{KT}{\xi}$ is the diffusion coefficient. The friction coefficient can be related with the radius (r) of the particle and the viscosity of the fluid (η) through the Stokes formula[**?**]:

$$\xi = 6\pi\eta r \tag{2}$$

And the diffusion coefficient can be calculated using the Stokes-Einstein equation[**?**]:

$$D_0 = \frac{KT}{\xi} \tag{3}$$

2 Methods

Given that the size of the analytes is much smaller than the size of the MNP we know that the time scale in which analytes found the first ligand is much smaller than the time scale in which analytes found the second one. Also we know that the binding energy between a biomarker and a ligand is quite big (13KT), consequently we can consider that bonds are irreversible. Using this two considerations we can decompose the problem in two parts: Binding of analytes with the first ligand, and binding of analytes with the second ligand (bonding between MNP).

2.1 Binding of analytes with the first ligand

The probability of finding a certain distribution of analytes on the MNP given an arbitrary number of MNP (N), of analytes (m) and of ligand per MNP (k) has been calculated both analytically and computationally.

The computational calculation has been made by representing the ensemble of all the MNP as an array 2D of dimensions (k x N), in which the element ij represents the ligand i of the MNP j. When the ligand ij is occupied the element i,j of the array is equal to 1, and when that ligand is unoccupied the element ij is 0. The simulation starts with all the elements of the array equal to 0, then we generate randomly pairs of integer numbers (i,j) uniformly distributed, i between 0 and k-1, and j between 0 and N-1, if the element ij is equal to 0, we change their value to a 1, by contrast when the element ij is equal to 1 we generate another pair. Repeating this process while the number of ones in the array is smaller than the number of biomarkers, we obtain a possible distribution of analytes. In order to get the probability that a certain distribution takes place, we repeated the algorithm described above one million times, and count the number of times that appear each distribution.

The mathematical analysis of the problem is quite complicated, the fully derivation can be found in the supplementary material, section 1. Briefly, we wanted to obtain a formula to predict the probability of finding an arbitrary number of MNP with all of them having a certain number (α) of ligands occupied, given an arbitrary number of MNP, analytes and ligands per MNP. To do that, we begin by calculating the probability of an arbitrary ensemble of MNP to have all of them α ligands occupied (on this first step we don't impose that all the particles with α ligands occupied are those of the ensemble). Then, using this calculation we derive the probability that only the particles in the ensemble have α ligands occupied. Finally, we multiply that probability by the number of possible ensembles of size $n\alpha$ that we can construct with N MNP to obtain the final result.

Once that we understand how is the distribution of the analytes through the MNP when they are only bonded to one ligand, we can begin the simulation of the formation of clusters.

2.2 Bounding of different MNP (cluster formation)

A Brownian Dynamics code has been written to simulate the MPNs movement and the formation of the clusters in the solution. The experimental radius of the MNP is 32 nm and the length of one ligands is $\sim$ 2 nm. In our simulation we used a system units in which the diameter of the MNP is 1, that means that we divided all the experimental distances by 64. In addition we used a system of units in which KT=D=1. The simulation was made in a periodic box of side 31, and the time step of the simulation was $1 \cdot 10^{-4}$ and we run the simulation for $1 \cdot 10^{8}$ time steps which is time enough for a MNP to go through the full box. In case that all the possible bonds were formed or that

during 10 million time steps no bonds were created we finish the simulation before completing all the time steps, because in both cases we can assume that the final configuration has been reached.

In the simulation the particles are not homogeneous spheres given that in some random points of the surface of the spheres there are ligands, for that reason it's important to have into account it's orientation. To solve this, we simulate each MNP as a regular icosahedron where each vertex is linked to the eleven left with an harmonic force. That force has an equilibrium distance equal to the separation between that vertex with the others in a regular icosahedron inscribed in a sphere of diameter 1. Each icosahedron can have 4 different kind of vertex. Vertex of type 0 are those that doesn't have a ligand, so they can't form a bond, vertex of type 1 have a ligand that is unoccupied and they can form a bond with a vertex of type 2 which are those that have a ligand that is occupied but is not forming a bond. Finally, vertex of type 3 are those that have a ligand that is linked to another one of a different MNP.

In order to improve the performance of our code we have used an algorithm of neighborhood search called linked cell list algorithm [**?**]. This algorithm consist in dividing the simulated space in smaller cubic cells of side l and only calculate the interaction forces among the particles of the same and the adjacent cells. The value of l must be equal or higher than a distance from which the interactions between the particles are negligible or zero. Using this algorithm the computational complexity of calculating the pairwise interaction among the MNP is reduced from $O(N^2)$ to $O(N)$.

The code was thought to be run in a GPU instead of a CPU because GPU are able to compute tons of operation in parallel (at the same time), which is very useful to have a great performance when the number of particles in the simulation is high. For that reason the full code is written to compute all the interactions in parallel.

2.2.1 Initialization of the simulation

To initialize the simulation we need to set its initial conditions, that means to set the initial position and orientation of all the MNP and to fix which vertex of each icosahedron have a ligand and which ligands are initially occupied.

The initial orientation of each MNP is obtained by multiplying all the position vectors of an icosahedron centered at the origin by a random rotation matrix. The position vectors of the vertex of an icosahedron centered at the origin are proportional to the set of vectors[**?**] $\{(\pm 1, \pm\varphi, 0),(\pm\varphi, 0, \pm 1),(0, \pm 1, \pm\varphi)\}$, where $\varphi = \frac{1+\sqrt{5}}{2}$ is the golden ratio. The random rotation matrix is generated as follows [**?**]: First we generate three random numbers: γ,α and sin(β. γ and α are the rotation angles around the z and x axis respectively and are uniformly distributed between -π and π, β is the rotation angle around the y axis and sin(β) is uniformly distributed between -1 and 1. Once that the three angles have been generated, we multiply each vector of the icosahedron by the rotation matrix $R(\alpha, \beta, \gamma) = R_z(\gamma)R_y(\beta)R_x(\alpha)$ where the matrix R_i are the rotation matrix around the i axis.

The initial position of each MNP is set by generating three random numbers uniformly distributed between -L/2 and L/2 which represent the coordinates (x,y,z) of the center of the icosahedron. Then we run a Monte Carlo simulation using a hard spheres potential between the MNP in which the interaction energy is infinity when the distance between two MNP is smaller than $2^{1/6} \cdot (2R)$ and 0 elsewhere. By running the Monte Carlo simulation we prevent two or more MNP from starting the simulation overlapped, which could cause a divergence of the simulation.

Finally, we set which vertex contains a ligand and which ligands are initially bonded to an analyte. In order to fix which vertex contains a ligand, we generate randomly as many different numbers between 0 and 11 as ligands per particle we want to have for each MNP, then among all the ligands we select randomly which ones are initially bonded to an analyte by repeating the algorithm explained in section 2.1.1.

2.2.2 Interactions of the simulation

Three different interactions have been considered in our model:

- The first interaction is the responsible of keeping all the vertex that represent a MNP together, forming an icosahedron inscribed in a sphere of radius R=32 nm. That interaction is an harmonic force between each vertex with the others where the equilibrium distance is equal to the distance between each pair of vertex in a regular icosahedron.

- The second interaction represents the bonding

force between two ligands that are linked to the same analyte, this force is also represented by an harmonic force with equilibrium length equal to 4 nm.

- Finally, we add another interaction that prevents two or more MNP from occupying the same space. We represent that interaction as a repulsive Lennard-Jones force:

$$\mathbf{F}(\mathbf{r}) = 4\varepsilon \left(12 \cdot \frac{\sigma^{12}}{r^{13}} - 6 \cdot \frac{\sigma^6}{r^7} \right) \mathbf{u_r} \tag{4}$$

Where σ is the diameter of the MNP. This force only acts when the distance (r) between a pair of MNP is smaller than $2^{1/6} \cdot \sigma$ (elsewhere the force is 0)

There are another force that we are neglecting in our model and that can cause differences between the experiments and the simulations. That force is caused by the magnetic interaction between the MNPs, and depends on how strong is the magnetic moment of the MNP. Comparing our results with the experimental measures we will check which is the range of parameters in which the approximation is valid.

2.2.3 Advance of a time step

Each time step begins by calculating the forces acting on each MNP, they are calculated by adding all the interactions that we have explained in the previous section. Then, we determine the next positions of all the vertex, to do that we use the equation (1), in which the term $\mathbf{d\tilde{r}}$ is calculated using the expression:

$$\mathbf{d\tilde{r}} = \sqrt{2 \cdot D_0 \cdot dt} \cdot \mathbf{u} \tag{5}$$

Where $\mathbf{u}$ is a vector of three random numbers generated with a Gaussian distribution of mean 0 and variance 1. Then we check if a new bond has been created, to do that we check the distance between vertex of type 1 and type 2. When the separation between a pair of them is smaller than 4 nm, a new bond is created.

2.2.4 Obtaining results

After running each simulation we get a list of all the pair of vertex that have been linked, and the position of all the MNP in the box, using the list of bonds, we can determine the number of clusters that have been created and which particles compose each cluster. That information serve us to estimate the hydrodynamic size (D_h) of each cluster, which is a measure of the size of a cluster and it's defined as the diameter of a sphere that diffuses at the same rate as the cluster that we are studying. The hydrodynamic radius (R_h) (half of the hydrodynamic size) can be calculated theoretically using the Kirkwood definition [**?**]:

$$R_h^{-1} = \left\langle r_{ij}^{-1} \right\rangle_{i \neq j} \tag{6}$$

Where r_{ij} are the pairwise distance. Another magnitude that we want to calculate is the polidispersity index (PDI) which is a measure of how disperses are the clusters in size, we calculate it as the standard deviation of the cluster sizes divided by their mean:

$$PDI = \frac{\sqrt{\langle D_h^2 \rangle - \langle D_h \rangle^2}}{\langle D_h \rangle} \tag{7}$$

Compactness of clusters can be estimated using the Flory exponent (ν), which is a number that relates the gyration radius (R_g) of clusters and the number of colloids that compose each of them, according to[**?**]:

$$R_g \propto N^{\nu} \tag{8}$$

It can be proved [**?**] that the Flory exponent of a straight rigid polymer is 1 while the Flory exponent of a globular cluster is 1/3. Any value between them correspond to a polymer that is more compact or more expanded depending on whether it is nearer than 1/3 or than 1. The gyration radius is a measure of the clusters size, similar but not equal to the hydrodynamic radius and it's defined as follows[**?**]:

$$R_g = \frac{1}{N+1} \cdot \sum_{i,j>i}^{N} \langle (r_i - r_j)^2 \rangle \tag{9}$$

The clusters distribution ($\mathcal{P}(i)$), which is the probability of an MNP to belong to a cluster of mass i, was also studied. It can be calculated according to:

$$\mathcal{P}(i) = \frac{i \cdot N(i)}{N_{tot}} \tag{10}$$

Where $N(i)$ is the number of clusters of mass i, and N_{tot} is the total number of particles.

2.2.5 Experimental measures

The experimental measures have been made using the dynamic light scattering (DLS) technique [**?**] which is a very useful method to measure the size of nanometric particles in a solution. Particle size is determined using a laser that illuminates the solution, when particles interact with the light of the laser they cause

fluctuations in the intensity of the scattered light. Using this fluctuations is possible to measure the speed and hence the diffusion coefficient (D_t) of particles. Finally, using the Stokes-Einstein equation [**?**] is possible to determine the hydrodynamic size (D_h) of the particle as:

$$D_h = \frac{k_b T}{3\pi\eta D_t} \tag{11}$$

Where k_b is the Boltzmann coefficient, T the temperature of the system and η the viscosity of the liquid.

3 Results and discussion

3.1 Binding of analytes with the first ligand

Figure **??** shows the probability distribution of finding a specific number of MNP with 1 (Figure **??**a) and 4 (Figure **??**b) ligands occupied (the figures corresponding to 0,2 and 3 ligands occupied can be found in the supplementary material, Figure S1), in a solution of 400 MNP, 4 ligands per MNP and different values of number of analytes. We are using the number of MNP and analytes instead of its concentration because in our model bonds are irreversible so equilibrium is always fully displaced to the state in which all the analytes are bonded to a ligand (if the number of analytes is equal or smaller than the number of ligands) so our results do not depend on the volume of the system.

Figure 2: *Probability that there is a certain number of MNP with exactly a) 1 ligand occupied and b) 4 ligands occupied, Results are obtained using both the simulation and the analytical formula. The number of MNP in all the curves is 400 and the number on ligands per MNP is 4.*

The different points in Figure **??** form gaussian-like curves whose center and amplitude depends on the number of analytes. In the case of 1 ligand occupied, when the number of analytes is small, the position of the center of the curves grows with the number of analytes until reaching a maximum position, then the position of the maximum turns back to 0, that's why in Fig **??**a the curve corresponding to 100 analytes is between the curves of 200 and 400 analytes. This behavior occurs because when there are few analytes most of the MNP are unoccupied, so the number of MNP with 1 ligands occupied is small, then when we increase the number of analytes there are more particles with 1 ligand occupied, but when the number of analytes is high, most of the particles have more than 1 ligands occupied, so the center of the curve goes back to 0. The same behavior was observed for the cases of 2 and 3 ligands per particle, the only change between that three cases is the positions of the centers and the number of analytes in which the center's position begin to decrease.

Analyzing the curve corresponding to 4 ligands we observe that increasing the number of analytes always increase the position of the center of the curves. This behavior continues until reaching the point where the ligands of all the MNP are occupied, in that point the distribution is a single point with probability 1 placed at the number of MNP of the simulation (in this case 400). This occurs because we are doing the simulations with particles that have 4 ligands, so, when we increase the number of analytes, slowly we are saturating the solution until occupying all the ligands (all the four ligands of the MNPs are occupied). The behavior observed for 0 ligands occupied is just the opposite, increasing the number of analytes moves the center of the curves to the left until reaching 0. Initially when the number of analytes is 0 the distribution is a single point with probability 1 and placed at the number of MNP of the simulation.

Points obtained using the simulation and using the analytical formula are almost indistinguishable for all the cases analyzed, the differences between them is always near a 1% which is of the same order of the difference between two simulations made with the same parameters. That means that both methods are valid to initialize the Brownian Dynamics simulation. We choose to use the simulation because computing the analytical formula for a high number of MNP and analytes is difficult because we have to compute the factorials of very high number which requires a very high computational cost.

3.2 Cluster formation

The formation of clusters was studied in first place maintaining constant the number of MNP and varying the concentration of analytes and the number of ligands per MNP. Then we keep constant the analytes concentration and the number of ligands per MNP and vary the MNP concentration.

Figure 3: *Mean hydrodynamic size of clusters as a function of the analytes concentration using a MNP concentration of* $7.64 \cdot 10^{14}$ *MNP/ml and a) 2 ligands/MNP, b)4 ligands/MPN, c) 8 ligands/MPN and d)12 ligands per MNP. In figures b,c,d results obtained are compared with experimental measures are obtained using the dynamic light scattering technique. The y-axis is plotted in logarithmic scale*

3.2.1 Maintaining constant the MNP concentration

Figure **??** shows the mean hydrodinamic size dependence on the analytes concentration when the MNP concentration is $7.64 \cdot 10^{14}$ MNP/ml and the number of ligands per MNP is 2, 4, 8 and 12 ligands per MNP.

In the case of 2 ligands/MNP when the analytes concentration is small the hydrodynamic size grows with the analytes concentration until reaching a maximum at the concentration of $\sim 1.25\mu M$, after that point the hydrodynamic size decreases symmetrically. This behavior occurs because when the analytes concentration is high most of the ligands are occupied, without the need of being linked to another, so the number of bonds created is small and so the hydrodynamic size. On the other hand when the number of analytes is small there are not enough of them to occupy all the ligands so the number of bonds is also small. The positions of the maximum can be calculated analytically using that the size is related with the maximum number of bonds that can occur in a simulation. That number is the minimum between the number of ligands which are initially occupied (n_{occ}) and those that are unoccupied (n_{unocc}), that means $n_{bonds} = min(n_{occ}, n_{unocc})$. That equation write in terms of the number of MNP (N), analytes (m) and ligands per MNP (k) is:

$$n_{bonds} = \begin{cases} 0 & if \quad m \geq Nk \\ Nk - m & if \quad Nk/2 < m < Nk \\ m & if \quad m < Nk/2 \end{cases} \tag{12}$$

From equation 12 it's easy to see that the maximum number of bonds occurs when the number of analytes is half of the number of ligands, in the case of 2 ligands/MNP this occurs for an analytes concentration of $1.27\mu M$ which is very close to the concentration where the maximum size is observed in the simulation that is $1.25\mu M$.

This same behavior is observed in the case of 4 ligands per particle, but in this case the maximum is reached for an analyte concentration of $\sim 2.5\mu M$ (again the half of the total number of ligands). Conversely, in the experimental measures the size of the clusters always grows with the analytes concentration, that is because in the experiments when the system is saturated with analytes, they can interact unspecifically (without the mediation of the ligands) with the MNP. The interaction energy of the unspecific interactions is much smaller than the binding energy of an analyte to a ligands, for that reason when the system is not saturated it's much more likely the interaction with ligands so the unspecific interactions do not cause variations between the experiment and the simulation. By contrast when the system is saturated unspecific interactions become relevant and are the responsible of creating new bonds between MNP. In the cases of 8 and 12 ligands per MNP we don't observe the saturation of the system in the computational simulations because it occurs at higher concentrations.

In the cases (4,8 and 12 ligands/MNP) in which we can compare the results of our simulation with experimental measures we can see that when the analytes concentration is small the agreement between the simulation and the experiment are very good, but for concentrations higher than $\sim 1.5\mu M$ the simulations predict a size much bigger than experiments. This occurs because in the simulation we are observing a phase transition to the gel phase (an unique cluster composed by all the MNP) which in the experiments do not occur. The explanation for this difference is that the MNP have a permanent magnetic dipole so in the experiments there are a magnetic interaction that we're not having into account in the simulation. Looking at the figures we can deduce that when the size of the clusters is not too big, the magnetic effects are negligible but when clusters became bigger the magnetic repulsion is strong enough to avoid the gelation of the system.

The influence of analytes concentration and ligands per MNP on the polidispersity index (PDI) was also studied, results obtained are shown in Figure **??**.

The behavior of the PDI in the case of 2 ligs/MNP is very different from the behavior of the other cases.

Figure 4: *PDI as function of the analytes concentration when the MNP concentration is* $7.64 \cdot 10^{14}$ *MNP/ml, and the number of analytes per MNP are a) 2 and b) 4 (green squares), 8 (orange squares) and 12 (blue squares)*

It presents a symmetric curve that has 2 absolute maximums centered in 0.5 and 2 μM and a local maximum centered in 1.25 μM. The shape of the curve is very striking because in general PDI is correlated with the mean size of clusters, however in this case that doesn't occur in two intervals: when the concentration of analytes is between 0.5 and 1 μM and between 1.5 and 2 μM. In that regions the mean size of the cluster grows because the number of small clusters decrease while the number and size of big clusters is more or less constant. That behavior causes an augmentation of the mean hydrodynamic size and a reduction of the dispersion.

By contrast in the other cases (4,8 and 12 ligs/MNP) PDI is always correlated with the mean size except when the system reaches the gel phase, where the PDI is 0 given that there is a unique cluster. The shape of the three curves is almost the same, the only difference is that for high concentrations the PDI corresponding to 4 ligs/MNP is not 0 because the system leaves the gel phase. When the analytes concentration is small the number of bonds that can be formed is limited only by the number of analytes and not by the number of ligands, that's why in that range the PDI is independent of the number of ligands/MNP (excepting the case of 2 ligs/MNP where the number of ligands is so small that is always important), by contrast when the analytes concentration is bigger ($\sim 2\mu M$) a higher number of ligands produces more variability and that causes a higher PDI.

The shape of the PDI can be better understood using the cluster distribution for each concentration. In Figure **??** is represented the probability of an MNP to belong to a cluster of mass i as a function of i for the cases of 2 ligands per MNP (Figure **??**a) and 4 ligands per MNP (Figure **??**b), the cases of 8 and 12 ligands per MNP are equal to the case of 4, so the analysis is the same (they can be found in the supplementary material, Figure S2).

In the case of 2 ligs/MNP when the analytes concentration is very small (0.1 μM) the most likely cluster is the formed by just 1 MNP, and the probability of higher clusters decreases very quickly to 0, when the analytes concentration to 0.5 μM there are bigger cluster that can appear and a shoulder begins to be formed around clusters of 2 particles. The fact that bigger clusters begin to appear is the responsible of the augmentation of the PDI in Figure **??**a at that concentration. When the concentration is $0.75\mu M$ the shoulder become a maximum because the number of clusters of 1 MNP is much smaller, this is the responsible of the decrease of the PDI. Finally for the concentration of $1.25\mu M$ the maximum is displaced to a mass a bit higher and clusters much higher have been formed, that's the reason of the relative maximum in the PDI. For higher concentrations of analytes, distributions turns back to the first one symmetrically.

In the case of 4 ligs/MNP probability distributions are always decreasing curves excepting the case in which the system is gelated. When we increase the analytes concentration big clusters are more likely, but the gelation occurs before we can observe the maximum observed in the case of two ligands per particle.

Red lines of the plots are a fit of the data to a negative binomial distribution which probability mass function is[**?**]:

$$\mathcal{P}(i) = \binom{i+r-2}{i-1} \cdot (1-p)^{i-1} \cdot p^r \qquad (13)$$

Where i=1,2,3... is the mass of the cluster, and p and r are parameters. p is a probability so it's value is between 0 and 1 and r is a positive integer number. Excepting the cases in which the system reaches the gel phase (Figure b, lower left panel) the fits are always very good, so we can conclude that cluster distribution follows a negative binomial distribution. When the distribution doesn't have a maximum the fitting curves have r=1, that is a particular case of the negative binomial distribution which is the geometric distribution [**?**]:

$$\mathcal{P}(i) = p \cdot (1-p)^{i-1} \qquad (14)$$

3.2.2 Maintaining constant the analytes concentration

Now we're going to analyze how varies size and PDI of cluster when we change the MNP concentration maintaining constant the number of ligands per MNP 4 and using two concentrations of analytes: 0.15 and 0.75 μM. Figure **??**a shows the variation of the hydrodynamic size with the MNP concentration. In both cases (0.15 and 0.75 μM) the qualitative behavior

Figure 5: *Cluster distribution for the cases of a) 2 ligs/MNP and b) 4 ligs/MNP and various analytes concentration, when the concentration of MNP is* $7.64 \cdot 10^{14}$ *MNP/ml. Blue and black dots are the points obtained in the simulation, and red lines are a fit of the data to a negative binomial distribution (equation 13). In figure a, the fitting parameters are: When [Analytes]=0.1* μM *r=1, p=0.8, when [Analytes]=0.75* μM *r=3 and p= 0.54, when [Analytes]=0.75* μM *r=3 and p=0.54 and when [Analytes]=1.25* μM *r=2 and p=0.11. In figure b, the fitting parameters are: When [Analytes]=0.1* μM *r=1, p=0.85, when [Analytes]=0.5* μM *r=1 and p= 0.4, when [Analytes]=1.0* μM *r=1 and p=0.1*

is very similar. When the MNP concentration is small, the size of the clusters is also small because the system is saturated and most ligands are already occupied without forming a bond. Initially the size grows very quickly with the MNP concentration due to the fact that increasing the number of MNP augment the number of ligands in the system and given that the system initially is saturated, this increases the number of bonds that can occur. If we keep increasing the MNP concentration there comes a concentration of MNP from which the size of the cluster begins to decrease, that is because the system is no longer saturated and increasing the number of MNP reduces the probabilities of various MNP with some ligands occupied to "meet" among them.

Figure 6: *a) Experimental (diamonds) and simulated (squares) hydrodinamic size as a function of the MNP concentration using a logarithmic scale in the y-axis. b) PDI obtained in the simulations as a function of the MNP concentration*

The position of the maximums can be predicted analytically using that they will occur when the number of bonds per particle is maximum, to find that maximum we will use the approximation that the number of bonds that are formed is the maximum number of bonds that can occur. Under that approximation the number of bonds per particle (n_{bonds}/N) is:

$$\frac{n_{bonds}}{N} = \begin{cases} 0 & if \quad Nk \leq m \\ \frac{Nk-m}{N} & if \quad m < Nk < 2m \\ \frac{m}{N} & if \quad Nk > 2m \end{cases} \tag{15}$$

Where N,m and k are the number of MNP, analytes and ligands per MNP respectively. The maximum of that function occurs when N=2m/k, in that point the function is not derivable because left and right limits of the derivative are different but the function is continuous and the sign of the derivative changes, so we know that it's a maximum. The predicted position when the analytes concentration is 0.15 μM is $[MNP] = 0.45 \cdot 10^{14}$ MNP/ml, while the obtained one is $[MNP] = 0.40 \cdot 10^{14}$ MNP/ml. When the analytes concentration is 0.75 μM the predicted position is $[MNP] = 2.26 \cdot 10^{14}$ and the obtained is $[MNP] = 2.50 \pm \cdot 10^{14}$. Given that we do the simulations at discrete values of the MNP concentration, in both cases the predicted position is inside the uncertainty of out results.

Our results of the sizes have been compared with the experimental measures. In the case of an analytes concentration of 0.15 μM the agreement of the results of the simulation with the all the experimental points is really good. Conversely, when the analytes concentration is 0.75 μM, when the MNP concentration is small experiment and simulation are very distant, but they approach when we increase the MNP concentration and the cluster size is reduced. Newly, we see that in the range of parameters where the size of the clusters is small the agreement is very good, but when clusters become bigger simulation and experiment separates. Again, this might be caused because of magnetic interactions between the clusters of MNP that in the experiment may avoid the formation of very big clusters, as occurs in the simulation.

In Figure **??**b it's plotted the PDI as a function of the MNP concentration for analytes concentrations of 0.15 and 0.75 μM. We can see that in both cases the dispersion is correlated with the size of clusters. In the case of 0.75μM PDI reaches a value of one for some MNP concentrations, that means that the final state is a very big cluster composed by the majority of the MNP and then a few small clusters. By contrast, in the case of 0.15 μM the PDI is always much smaller than 1, so there are not a huge difference between the sizes of clusters. That supports the theory that magnetic force is the cause of the differences between the experiments and the simulation because in the experiments it avoids the formation of a cluster composed by majority of the MNP.

3.2.3 Compactness of the clusters. Flory exponent

Using all the clusters that have been formed in every simulation we have made the Figure **??** that shows

the mean gyration radius of the clusters dependence on the number of particles that compose each cluster. Gyration radius can vary from one cluster to another that has the same number on MNP, for that reason we have represented the mean gyration radius of all the clusters with the same number of MNP.

Figure 7: *Gyration radius as a function of the number of particles that compose each cluster in logarithmic scale, for the cases of a) 2 ligands/MNP, b) 4 ligands/MNP, c) 8 ligands/MNP and d) 12 ligands/MNP. Each point of the figures have been obtained by averaging the gyration radius of all the clusters with the same number of MNP. The errorbars are the quadratic mean error. e) Mean number of neighbors that a MNP have in a cluster as a function of the mass of the cluster*

As predicted by Flory's theory **[?]** the points of all the figures are aligned forming a straight line when we use a logarithmic scale in both axis, Flory's coefficient (ν) is equal to the slope of the straight lines. By doing a least square fit we obtain the Flory coefficients of each case. They are collected in table **??**.

Ligs./MNP	ν
2	0.71 ± 0.02
4	0.61 ± 0.02
8	0.61 ± 0.02
12	0.62 ± 0.02

Table 1: *Flory exponents for 2,4,8 and 12 ligands/MNP*

The Flory exponent when the number of ligands per MNP is 2 is $\nu \sim 0.71$ which correspond to a very expanded chain, as we expected this value is the highest of all the cases because of the fact that in this case each MNP can be only linked to a maximum of two MNP while others can be linked to more. By contrast, in all other cases the Flory exponent is almost the same ($\nu \sim 0.61$) despite the fact that the number of links that can be formed changes in each case. This is because even if a MNP with 12 or 8 ligands/MNP in principle can form 12 or 8 bonds respectively, this is very unlikely to happen because when a particle (A) is bonded to another 3 or 4 and another one approaches to it, that MNP will find first any of the particles that are linked to A, than A. In fact, as it can be seen in Figure **??**e the mean number of neighbors that each MNP have in a cluster is between 2 and 3 for all the cases except $Ligs./MNP = 2$, where given that the maximum number of bonds that particles can form is 2, the mean is a little bit smaller.

4 Conclusions

In the present work we have performed a computational study of the formation of clusters of MNP after their specific interactions with a biomarker.

By comparing the results of the simulation with the experimental measures we have seen that there are two interactions apart from the Lennard-Jones repulsion between the MNP and the interaction between ligands and analytes that are very important in some ranges of parameters. The first of them is the unspecific interaction between the MNP and the biomarkers which is the responsible than when the number of analytes is very big (related with the number of ligands) the size of the clusters keeps growing with the analytes concentration even if the system were saturated of analytes. The other interaction is the magnetic force among the clusters of MNP which is the responsible of avoiding the formation of very big clusters and the gelation of the system in the experiments when in the simulations it happens.

In addition we have seen that when we vary the concentration of MNP there are an optimal concentration for which the cluster size reaches a maximum, and we have proved that the maximum appears when the number of bonds per particle that can be formed is maximum.

Also we have seen that excepting the case when the number of ligands per MNP is 2, the PDI is always correlated with the size of the clusters, also that when the cluster size is big increasing the number of ligands per MNP increase the PDI. In addition, we have studied the cluster mass distribution and we have seen that in all the cases where the system is not gelated, that distribution fits very well to a negative binomial distribution.

Finally, we have seen that when the number of ligands per MNP is equal or higher than 4, that magnitude do not affect the compactness of the clusters because is very unlikely that an MNP forms more than 3 bonds. The Flory exponent when the number of ligands per MNP is equal or higher than 4 is ~ 0.61 which is a expanded chain. In the case of 2 ligands per MNP the Flory exponent is ~ 0.71 which is a chain more expanded than in the other cases.

References

[1] Mejías et al. Biomaterials 32, 2938 (2011)

[2] Klauss Maier-Hauff et al. J Neurooncol 103,317 (2011) DOI 10.1007/s11060-010-0389-0

[3] Rocha-Santos Trends in Analytical Chemistry, 62,28 (2014)

[4] Fornara et al. Nanoletters 2008 DOI 10.1021/nl8022498; Donolato et al Anal. Chem. 87, 1622 (2015); Wang et al. Sci Rep. 4, 5716 (2014); Perez et al. J. Am. Chem Soc 125, 10192 (2003)

[5] https://www.ntsol.es/

[6] Paul Langevin's 1908 paper "On the Theory of Brownian Motion" ["Sur la théorie du mouvement brownien," C. R. Acad. Sci. (Paris) 146, 530–533 (1908)]

[7] Laidler, Keith J.; Meiser, John H. (1982). Physical Chemistry. Benjamin/Cummings. p. 833. ISBN 0-8053-5682-7.

[8] X. Bian, C. Kim and G. Karniadakis, Soft Matter, 2016, DOI: 10.1039/C6SM01153E.

[9] Weisstein, Eric W. "Regular Icosahedron." From MathWorld–A Wolfram Web Resource. https://mathworld.wolfram.com/Regular Icosahedron.html

[10] Xavier Perez-Sala, Laura Igual, Sergio Escalera, Cecilio Angulo. (2012). Uniform Sampling of Rotations for Discrete and Continuous Learning of 2D Shape Models

[11] Matthew Dobson and Ian Fox and Alexandra Saracino (2014). Cell List Algorithms for Nonequilibrium Molecular Dynamics

[12] . Kirkwood, J. G. (1954). The general theory of irreversible processes in solutions of macromolecules. J. Polym. Sci., Polym. Phys. Ed. 12:1–14.

[13] Paul J. Flory. The Configuration of Real Polymer Chains. J. Chem. Phys. 17, 303 (1949)

[14] Marshall Fixman. Radius of Gyration of Polymer Chains. J. Chem. Phys. 36, 306 (1962)

[15] Jörg Stetefeld, Sean A. McKenna and Trushar R. Patelcorresponding. Dynamic light scattering: a practical guide and applications in biomedical sciences. Biophys Rev. 2016 Dec; 8(4): 409–427.

[16] M. Rubinstein and R. H. Colby, Polymer Physics, Oxford Univ. Press, 2003.

[17] Cook, John D. "Notes on the Negative Binomial Distribution"

[18] Weisstein, Eric W. "Geometric Distribution." From MathWorld–A Wolfram Web Resource. https://mathworld.wolfram.com/Geometric Distribution.html

5 Supplementary material

5.1 Mathematical calculation of the probability of finding an arbitrary number of MNP, each of them with the same number of ligands occupied.

Given a number N of MNP, m of analytes and k of ligands per MNP we will obtain an equation to predict the probability of finding exactly a number n_α of MNP with each of them having α ligands occupied after preparing a solution with that parameters. The formula is derived in three steps: first, we derive the probability $\tilde{P}(n_\alpha^+)$ of finding an arbitrary ensemble $\{n_\alpha\}$ of n_α MNP with each particle in the ensemble having α ligands occupied, then we will obtain the probability $\widetilde{P}(n_\alpha)$ of only the n_α particles in the ensemble have α ligands occupied, finally, we derive the probability of finding exactly n_α MNP with α ligands occupied.

5.1.1 Probability of an arbitrary ensemble of MNP having α ligands occupied

We will calculate the probability $\tilde{P}(n_\alpha^+)$ using a tree diagram in which each node correspond to a new bond between an analyte and a ligand and has two branches: one corresponding to the case that the ligand belongs to $\{n_\alpha\}$ and the other if not. Note that there are a lot of paths in the diagram which arrives to a possible configuration, but all of them have the same probability given that the number of ligands occupied inside $\{n_\alpha\}$ is always the same and that each bonding is an independent event. Because of that we will obtain the probability of taking place one path and then we will multiply that probability by the number of possible paths. In each node the probability of occupying a ligand inside $\{n_\alpha\}$ is the number of free ligands in the set divided by the total number of ligands, on the other hand, the probability of occupying a ligand outside $\{n_\alpha\}$ is the number of free ligands outside the ensemble divided by the total number of free ligands, given that the denominator is the same in both cases and decreases one in each node from the total number of ligands (Nk) to the final number of free ligands (Nk-m), we know that one factor of the probability is:

$$T1 = \frac{1}{kN} \cdot \frac{1}{kN-1} \cdot ... \cdot \frac{1}{kN-m+1} = \frac{(Nk-m)!}{(Nk)!} \tag{16}$$

Now, we will calculate the factor corresponding to the occupation of ligands outside the ensemble. Given that the total number of bonds is m and the number of bonds inside is $\{n_\alpha\}$ is αn_α, the number of occupations outside the ensemble will be $m - \alpha n_\alpha$, after each occupation the number of available ligands (favorable cases) is reduced by one starting from $(N - n_\alpha)k$, and arriving to $(N - n_\alpha)k - m - \alpha n_\alpha$, so the factor associated to the occupation of ligands outside $\{n_\alpha\}$ is:

$$T_2 = \frac{[k(N-n_\alpha)]!}{[k(N-n_\alpha)-(m-\alpha n_\alpha)]!} \tag{17}$$

The term associated to the binding of ligands inside $\{n_\alpha\}$ is similar to the previous term in the sens that after each link the number of favorable cases is reduced by one, but in this case we have to exclude the cases in which one MNP has more ligands occupied than α. To do that we will begin by calculating the probability of occupying an arbitrary set of ligands of each MNP, for example in the case of $\alpha = 2$ and k=3, we impose that the ligands occupied were always the ligand A and B and not C. Doing that, the third factor can be calculated equal than the second factor, knowing that the initial number of ligands available is $\alpha n\alpha$ and the last term will be 1, so the third factor is:

$$T_3 = (\alpha n\alpha)! \tag{18}$$

To generalize the last factor to any group of α ligand per MNP we multiply by the number of groups of α ligands that we can make with k ligands, we have to make this multiplications so many times as MNP are in $\{n_\alpha\}$ so the fourth term is:

$$T_4 = \binom{n_\alpha}{k}^{n_\alpha} \tag{19}$$

With this four factors we have the probability of an valid path in the tree diagram, to obtain the total probability we have to multiply by the number of possible paths in the diagram, which is the number of combinations of $\alpha \cdot n_\alpha$ nodes that we can make with a total of m nodes, that is:

$$T_5 = \binom{\alpha n_\alpha}{m} \tag{20}$$

Joining all the terms, the probability $P(n_\alpha^+)$ of an arbitrary ensemble of n_α MNP having α ligands

occupied is:

$$\widetilde{P}(n_\alpha^+) = \binom{k}{\alpha}^{n_\alpha} \cdot \binom{m}{n_\alpha} \cdot (\alpha n_\alpha)! \cdot \\ \cdot \frac{[k(N-n_\alpha)]!}{[k(N-n_\alpha)-(2m-\alpha n_\alpha)]!} \cdot \frac{(Nk-m)!}{(Nk)!} \tag{21}$$

5.1.2 Probability of only an arbitrary ensemble of MNP having α ligands occupied

The equation derived in the last section includes the probability of finding all the ensembles with $n_\alpha + 1$, $n_\alpha + 2$... MNP which contains the set $\{n_\alpha\}$. The only case in which this don't occur is when n_α is the maximum number of MNP which can have α analytes bonded (n_α^{max}) given a set of parameters, in this case $\widetilde{P}(n_\alpha^+) = \widetilde{P}(n_\alpha)$. It can be proved that n_α^{max} is:

$$n_\alpha^{max} = \begin{cases} floor\left(\frac{m}{\alpha}\right) & if \quad m \leq N\alpha \\ floor\left(N - \frac{m-N\alpha}{k-\alpha}\right) & if \quad m > N\alpha \end{cases} \tag{22}$$

If we know $\widetilde{P}(n_\alpha^{max})$, we can determinate $\widetilde{P}(n_\alpha^{max}-1)$ knowing that, there are $N-(n_\alpha^{max}-1)$ sets that contain $\{n_\alpha^{max}-1\}$ and have n_α^{max} MNP with α analytes:

$$\widetilde{P}(n_\alpha^{max}-1) = \widetilde{P}(n_\alpha^{max}-1)^+ - [N- \\ -(n_\alpha^{max}-1)] \cdot \widetilde{P}(n_\alpha^{max}) \tag{23}$$

In the same way, we can calculate $\widetilde{P}(n_\alpha^{max}-2)$, using that there are $N-(n_\alpha^{max}-2)$ sets that have $n_\alpha^{max}-1$ MNP and $\binom{N-(n_\alpha^{max}-2)}{2}$ sets with n_α^{max} MNP. Following this procedure the probability for an arbitrary α is:

$$\widetilde{P}(n_\alpha) = \widetilde{P}(n_\alpha^+) - \sum_{n=n_\alpha+1}^{n_\alpha^{max}} \binom{N-n_\alpha}{n-n_\alpha} \cdot \widetilde{P}(n) \tag{24}$$

This calculation is easy to do computationally, we begin by calculating $\widetilde{P}(n_\alpha^{max})$ and use that value to determine the others using the already calculated values to calculate the following.

5.1.3 Probability of finding exactly n_α MNP with α ligand occupied

Finally, we just need to multiply the probability obtained in the last section by the total number of sets of n_α MNP that we can construct with a total of N MNP to obtain the total probability, that is:

$$P(n_\alpha) = \binom{N}{n_\alpha} \widetilde{P}(n_\alpha) \tag{25}$$

5.2 Figures

Figure 8: *a) Experimental (diamonds) and simulated (squares) hydrodinamic size as a function of the MNP concentration using a logarithmic scale in the y-axis. b) PDI obtained in the simulations as a function of the MNP concentration*

Figure 9: *Cluster distribution for the cases of a) 8 ligs/MNP and b) 12 ligs/MNP and various analytes concentration, when the concentration of MNP is* $7.64 \cdot 10^{14}$ *MNP/ml. Blue and black dots are the points obtained in the simulation, and red lines are a fit of the data to a negative binomial distribution. In figure a, the fitting parameters are: When [Analytes]=0.1 µM r=1, p=0.85, when [Analytes]=0.5 µM r=1 and p= 0.42, when [Analytes]=1.0 µM r=1 and p=0.1. In figure b, the fitting parameters are: When [Analytes]=0.1 µM r=1, p=0.85, when [Analytes]=0.5 µM r=1 and p= 0.4, when [Analytes]=1.0 µM r=1 and p=0.1*